

Kathrin Steyer

Zwischen theoretischer Modellierung und praxisnaher Anwendung. Zur korpusgesteuerten Beschreibung usueller Wortverbindungen

Abstract

This paper¹ shows how a corpus-driven approach leads to a new perspective on central issues of phraseology and on lexicographical applications. It argues that a data-driven pattern search (applying statistical methods), an a posteriori interpretation of the data and a user oriented documentation of the usage of multi-word units (e. g. in lexicographical articles) constitute a step-by-step process where each step has its own informational value and usefulness. The description of multi-word units (*Usuelle Wortverbindungen*) presented in this paper focuses on the second step, the high quality analysis and interpretation of collocation data, exemplified by the fields of multi-word units centered around the word forms *Idee/Ideen/(idea/ideas)*.²

1. Einleitung

Die Termini »Korpora« und »Korpusbasiertheit« spielen mittlerweile bei einer wachsenden Zahl von Modellen eine zentrale Rolle. Damit einhergehend nimmt auch die Heterogenität der Auffassungen, Methoden, Erklärungsansätze und Nutzungsperspektiven zu. Je disparater der Objektbereich und das Methodenspektrum sind, desto dringlicher scheint eine begleitende methodische Reflexion, vor allem in Hinblick darauf, wie mit einer ganz bestimmten korpusanalytischen Methode entdeckte sprachliche Phänomene zu interpretieren sind.

Nach wie vor wird oft zu wenig explizit gemacht, auf welche Weise man die Sprachdaten befragt und interpretiert sowie Verallgemeinerungen vornimmt. Eine Unterscheidung betrifft die Art, wie man sich den Korpusdaten »nähert«, ob man also »korpusbasiert« oder »korpusgesteuert« vorgeht.

Korpusbasiertes Herangehen arbeitet mit im Vorhinein aufgestellten Hypothesen und Fragen an das Korpus. Es dient primär der Überprüfung und gegebenenfalls Revidierung, in jedem Falle aber der Ausdifferenzierung und Erweiterung bereits vorhandenen Wissens über sprachlichen Gebrauch. **Korpusgesteuertes** Herangehen ist erklärtermaßen induktiv und erst auf der Suche nach rekurrenten sprachlichen Strukturen. Es versucht, das Wissen aus den beobachteten Sprachdaten zu generieren und Generalisierungen *a posteriori* vorzu-

¹ Ich bedanke mich bei Annelen Brunner für die kritische Diskussion meines Manuskripts. Von Peter Dur̃co habe ich wichtige Anregungen und Impulse erhalten, die vor allem in den 4. Punkt eingeflossen sind.

² Dieser Beitrag ist in wesentlichen Teilen bereits 2006/2007 erarbeitet worden und dokumentiert den damaligen Forschungsstand.

nehmen.³ Beide methodischen Wege haben ihre Berechtigung. Kein Paradigma kann derzeit den Alleinvertretungsanspruch erheben und wird es vielleicht auch nie können. In diesem Beitrag werde ich ein korpusgesteuertes Beschreibungsmodell für Wortverbindungen entwickeln, das auf der linguistischen Interpretation von mathematisch-statistisch ermittelten Kookkurrenzdaten basiert.

Drei Prämissen bilden das methodische Kredo des in diesem Beitrag vorgeschlagenen Ansatzes:

- I. Korpusgesteuerte qualitative Sprachbeschreibung unterscheidet zwar zwischen »besseren« (typischen) oder »schlechteren« (weniger typischen) Kandidaten in Hinblick auf den sprachlichen Usus, aber nicht in Hinblick auf linguistisch-kategoriale Einordnungen. Es gibt kein Zentrum und keine Peripherie in der linguistischen Relevanzsetzung. Damit befindet die korpusgesteuerte Sprachbeschreibung nicht *a priori* darüber, welche Kandidaten beschreibungswürdiger als andere sind. Sie strukturiert die vorgefundenen Daten vielmehr *a posteriori*.⁴
- II. Korpusgesteuerte qualitative Sprachbeschreibung ist primär pragmatisch ausgerichtet. Der Topos, dass Bedeutung Gebrauch ist, manifestiert sich in diesem methodischen Paradigma auf besonders ausgeprägte Weise. Funktionales wird aus der Nische subjektiver Perspektiven auf einzelne sprachliche Akte herausgeholt und durch die Möglichkeit, massenhafte Gebrauchsmuster zu erkennen, »kollektiviert« und damit besser generalisierbar.
- III. Korpusgesteuerte qualitative Sprachbeschreibung lässt die Daten – so weit es geht – zunächst für sich selbst sprechen und zieht die ordnenden Fäden im Hintergrund. Dem Betrachter soll ein strukturierter, dabei aber möglichst unverfälschter Blick ins Korpusuniversum und damit auf authentische Sprache ermöglicht werden. Im Ergebnis soll er die realen Sprachdaten in seinen eigenen Interpretationsrahmen einbetten können.

Arbeitet man als Linguist konsequent korpusgesteuert, wird man mit neuen Fragen und aber auch neuen Problemen der phraseologischen Beschreibung konfrontiert. Neue interpretative Herausforderungen entstehen z. B. durch die Quantität und die Heterogenität der beobachteten Sprachdaten. Man sieht nicht nur einfach mehr als früher. Man hat auch eine andere Sicht auf die Dinge und erkennt neue Zusammenhänge, Vernetzungen, Hierarchien. So können sich beispielsweise bisher als prominent und daher auch lexikografisch beschreibungswürdig angesehene Wortverbindungen als weniger relevant erweisen.

³ Verwiesen sei hier auf das Gesamtwerk von John Sinclair (stellvertretend Sinclair 1991). Zu methodischen Ansätzen in diesem Paradigma vgl. u. a. Tognini-Bonelli (2001), Hanks (2004); Belica/Steyer (i. Dr.), Perkuhn (2007), Steyer/Lauer (2007), Brunner/Steyer (2007). Zu weiteren Aspekten aktueller korpuslinguistischer Forschung vgl. auch Näf; Duffner (2006), Lemnitzer/Zinsmeister (2006).

⁴ Eine Problematisierung mit so genannten Zentrum-Peripherie-Modellen, in Bezug auf Phraseologismen (z. B. Fleischer 1997), nimmt bereits Feilke vor (2004: 45ff.). Donalies hat sich mit der Problematik der so genannten »Ränder« oder der marginalen sprachlichen Phänomene in Bezug auf die Morphologie in einem der Prämissen I entsprechenden Sinne auseinandergesetzt (Donalies 2007).

Es kann dagegen andere Kandidaten geben, die sich durch eine viel größere Auffälligkeit auszeichnen, als bis dato zu erkennen war. Tradierte Phraseologismen stellen sich dann als gar nicht so unikal und einzigartig heraus. Es geraten neue Vertreter ähnlicher Art, andere Realisierungen gleicher Grundmuster, Bildungsmuster höherer Ordnung in den Blick. Sie sind weder immer irregulär noch immer regelkonform. Sie sind aber immer gebrauchsgelitet erklärbar und damit in jedem Fall pragmatisch restringiert (vgl. Feilke 1996). Viele dieser Mehrwort-Vorkommen mussten sich unseren Analysen im vortechnologischen Zeitalter noch verschließen.

Drei Grundsatzfragen der Phraseologieforschung und der Phraseografie, die die Fachdiskussion bestimmten und immer noch bestimmen, werde ich aus korpusgesteuerter Sicht betrachten:

- Welche Einheiten sind die Bezugsgrößen für eine qualitative Beschreibung und Dokumentation üblicher Wortverbindungen einer Sprache?
- Welcher Art von Varianz unterliegen diese Einheiten und wie sind sie zu beschreiben?
- Wie ist das Verhältnis von korpusgesteuert erarbeiteten Ergebnissen als Beitrag zur linguistischen Theoriebildung einerseits und einer möglichst praxisnahen Verarbeitung dieser Resultate andererseits zu bestimmen?

2. Die Eingrenzung des Beschreibungsgegenstands *usuelle Wortverbindungen*

2.1 Das »subjektive« Kriterium der Idiomatizität

In der Regel versucht man, die Frage nach den genuinen Beschreibungsobjekten der Phraseologie und Phraseografie auf dem Weg der Rekonstruktion der besonderen Restriktionen und Beschränkungen zu lösen, denen eine Folge von mindestens zwei sprachlichen Einheiten über die normalen morphosyntaktischen und semantischen Regeln hinausgehend oder abweichend von ihnen unterliegt (Burger 2007: 11 ff.). Besondere Restriktionsbedingungen soll hier bedeuten, dass die Komponenten nicht beliebig austauschbar sind, ohne die Übersummativität ihres Kovorkommens zu verändern oder zu verlieren. Es geht darum zu klären, was viele Phraseologismen von freien Wortverbindungen abhebt (Burger 2007: 31).⁵ Ein nach wie vor prominentes Kriterium zur Abgrenzung von Phraseologismen und freien Wortverbindungen ist die semantische Idiomatizität. Eine Wortverbindung gilt oft als umso interessanter und damit beschreibungswürdiger, je weniger transparent, also je idiomatischer sie ist oder zu sein scheint. Deshalb wird das Kriterium der Idiomatizität auch häufig bemüht, wenn es um die Aufnahme und die Beschreibung von Wortverbindungen in Wörterbüchern oder Lehrwerken/Lernmaterialien geht. Kann es jedoch ein objektives Kriterium für die Bestimmung des Idiomatizitätsgrades einer Wortverbindung geben? Die nicht-kompositionelle (sprich idiomatische) Qualität einer Wortverbindung ist ja nicht per se gegeben, ihr sozusagen inhärent. Sie kann nur abgeleitet werden auf der Basis des zu Grunde gelegten Bedeutungsspektrums ihrer Komponenten. Kompetenzbasierte Lesartenzuschreibungen können je nach semantischer Schule oder subjektiver Auffassung eines

⁵ Burgers *Einführung in die Phraseologie*, mittlerweile in der 3. Auflage erschienen (2007), und die HSK-Bände zur Phraseologie (vgl. Burger/Dobrovol'skij/Kühn/Norricks 2007) können als die Standard- und Referenzwerke der modernen phraseologischen Forschung gelten.

Lexikografen sehr unterschiedlich ausfallen. Die Korpusanalyse hilft, den idiomatischen Status objektiver zu beurteilen, wie die folgenden Beispiele zeigen (vgl. auch Brunner/ Steyer 2007):

(1a) *sprüht vor Ideen*

Die Wortverbindung drückt aus, dass jemand ein großes Maß (und manchmal Übermaß) an Gedanken und Einfällen besitzt, was sich auch durch eine ausgeprägte Emotionalität äußern kann. Geht man von der Bedeutungsdefinition des *Duden* – *Das große Wörterbuch der deutschen Sprache* (GWDS 2000) für das Verb *sprühen* aus, müsste es sich hier um eine teil-idiomatische Wortverbindung handeln. Im GWDS werden die entsprechenden Kombinationen mit *sprühen* als übertragen gekennzeichnet, ohne eine eigene Lesart zu sein.⁶ Die automatische Umgebungsanalyse des Verbs *sprühen* ergibt nun aber ein reiches nominales Feld von solchen Partnerwörtern wie *Witz/Tatendrang/Lebensfreude/Energie/Optimismus/Ehrgeiz/Zuversicht* usw. Sie sind Komponenten des verbalen Wortverbindungsmodells *vor etw. sprühen*. Man könnte also durchaus eine normale Lesart für *sprühen* in Verbindung mit der Präposition *vor* im Sinne von 'ein großes Maß/Übermaß an etwas hervorbringen/sich in einem solchen Zustand befinden' ansetzen.

(1b) *fixe Idee*

Die Wortverbindung bezeichnet einen als verrückt, nicht Erfolg versprechend, sinnlos angesehenen Gedanken/Einfall, der oft fast zwanghaft verfolgt wird. Sprecher verwenden diese Wortverbindung zumeist nur auf andere Personen, nicht aber auf sich selbst bezogen. Es gibt keine weiteren Partnerwörter, mit denen das Adjektiv *fix* im Sinne von 'verrückt, spinnert, nicht Erfolg versprechend, zwanghaft' verwendet wird. Damit ist *fix* relativ unikal, d.h. nur in dieser speziellen Wortverbindung in dieser Bedeutung vorkommend. Die Wortverbindung kann also als teil-idiomatisch interpretiert werden.

(1c) *Gute Idee*

Diese Wortverbindung scheint auf den ersten Blick semantisch transparent und daher eher zu vernachlässigen. Die orthografische Variante des Adjektivs *gut* in Großschreibung *Gut* ist jedoch ein gesonderter signifikanter Partner von *Idee* (neben *gut*) und indiziert eine Auffälligkeit im Gebrauch. Die Analyse der Kontextzeilen zeigt, dass es sich um eine, in vielen Fällen in den Korpora sogar als explizit mündlich markierte, kommunikative Formel handelt (z.B. wie »Gute Idee«, kommentierte sogar X; »Stimmt! Gute Idee!« oder *Gute Idee übrigens*). Die Wortverbindung ist hier also, obwohl nicht-idiomatisch, pragmatisch restringiert.

Obwohl alle drei Fälle eine graduell unterschiedliche Idiomatizität aufweisen, sind sie aus korpusgesteuerter Sicht völlig gleichberechtigt zu betrachten.⁷ Ihre jeweiligen Komponenten *Idee-sprüht*, *Idee-fixe*, *Idee-Gut* weisen eine statistisch auffällige Affinität zueinander auf. Diese Affinität resultiert jedoch **allein** aus dem rekurrenten Gebrauch genau dieser korrelierenden sprachlichen Formen im Korpus. Auch für die folgenden Beispiele aus dem

⁶ »**sprühen** <sw. V.> [...]: **1. a)** in vielen kleinen, fein verteilten Tröpfchen, in zerstäubter Form an eine bestimmte Stelle gelangen lassen [...] Ü [...] ihre Augen sprühten [vor Freude]; er ließ seinen Geist, Witz s.; der Redner sprühte von Ideen, vor Witz; <häufig im 1. Part.:> sein sprühender (reger, immer neue Ideen hervorbringender) Geist; in sprühender (ausgelassener) Laune; ein sprühendes (besonders lebhaftes) Temperament [...]« © 2000 Dudenverlag (GWDS 2000).

⁷ »Die syntaktisch-semantische Konstituenz des Ausdrucks ist nicht die Ursache seiner idiomatischen Prägung« (Feilke 2004: 49).

Wortverbindungsfeld zum Adjektiv *geistig* erklären nicht sprachsystematische Gründe die Festigkeit der Wortverbindungen: *Geistige Heimat* weist eine stärkere Signifikanz auf als *geistige Herkunft*, *geistige Frische* als *guter geistiger Zustand*, *Kapital verspielen* als *Potenzial verspielen*. Die objektive Tatsache einer Gebrauchsrekurrenz ist für all diese Vorkommen zunächst einfach zu konstatieren, unabhängig davon, welche Post-Hoc-Erklärungsmuster für die Art der Beschränkung, ob syntaktisch, semantisch oder pragmatisch, anzusetzen sein könnten.

2.2 Das »objektive« Kriterium der Gebrauchsrekurrenz

Gebrauchsrekurrenz lässt sich nicht kompetenzbasiert herleiten. Sie kann erst durch mathematisch-statistische Verfahren auf der Basis sprachlicher Massendaten erkannt werden. Die primäre empirische Grundlage, auf der das Beschreibungsmodell der usuellen Wortverbindungen fußt, bilden die Ergebnisse der statistischen Kookkurrenzanalyse (Belica 1995). Diese Methode ermittelt, welche lexikalischen Einheiten⁸ ungewöhnlich häufig in der Umgebung eines Bezugswortes zu finden sind und somit einen statistisch signifikanten Kohäsionsgrad zu ihm aufweisen. Belegstellen aus dem Korpus werden nach diesem Merkmal gruppiert.⁹ Linguisten können die Ergebnisse u. a. in Form von Kookkurrenzlisten auswerten, in denen die Relationen der Kookkurrenzpartner und ihrer entsprechenden Originaltextausschnitte im Korpus in Cluster zusammengefasst und mit statistischen Angaben versehen sind.

Abbildung 1. zeigt Ausschnitte solcher Kookkurrenzlisten zum Bezugswort *Idee* in einer vereinfachten Grafik; zum einen die hierarchische Beziehung zwischen dem Bezugswort, dem primären Kookkurrenzpartner und weiteren sekundären Kookkurrenzpartnern (fett gedruckt), die dieses Kovorkommen weiter spezifizieren; zum anderen die typischen kontextuellen Einbettungen der kookkurrierenden lexikalischen Einheiten innerhalb eines Satzes (»syntagmatische Muster«) (kursiv gedruckt).

□ **gekommen** LLR: 21711 Gesamttreffer: 3553

- **gekommen nie wäre** 80% *wäre [...] nie auf die Idee [...] gekommen* Treffer: 125
- **gekommen nie sei** 57% *sei [man] nie auf die Idee gekommen* Treffer: 14
- **gekommen nie** 92% *wäre nie auf die Idee [...] gekommen* Treffer: 297
- **gekommen wäre** 82% *wäre nie auf die Idee [...] gekommen* Treffer: 468
- **gekommen sei** 69% *sei ihm auf die Idee [...] gekommen* Treffer: 445
- **gekommen** 97% *auf die Idee [...] gekommen* Treffer: 3553

⁸ Ich argumentiere im Folgenden immer auf der Ebene lexikalischer Einheiten, obwohl man exakterweise von »Zeichenketten« sprechen müsste, da die Kookkurrenzanalyse prinzipiell auf Zeichenketten jeglicher Art aufsetzen kann (im Prinzip könnten es auch Metadaten und andere Symbole sein). In der Regel sind es jedoch Wörter und diese sollen uns hier auch nur interessieren.

⁹ Die Kookkurrenzanalyse fasst Originaltextausschnitte zu hierarchischen Gruppen, den [...] Clustern, zusammen, z.B. aufgrund eines gleichen bzw. ähnlichen Kontextverhaltens eines zentralen Trefferobjekts.« (Perkuhn 2007: 482). Zum linguistischen Erklärungswert dieser Methode auch für phraseologische Fragestellungen vgl. Steyer (2000, 2003, 2004b) und Steyer/Lauer (2007).

□ **kam** LLR: 16208 Gesamttreffer: 5870

- **kam rettende** 80% *kam ... die rettende Idee* Treffer: 62
- **kam verhängnisvolle** 81% *kam ihm auf eine |die verhängnisvolle Idee* Treffer: 11
- **kam ausgerechnet** 86% *auf die Idee kam [...] ausgerechnet Uwe Ochsenknecht* Treffer: 46
- **kam** 54% *kam [... auf |die] Idee* Treffer: 5870

□ **kommen** LLR: 10577 Gesamttreffer: 2685

- **kommen nie könnte** 100% *nie ... Idee kommen könnte* Treffer: 1
- **kommen nie** 96% *würde nie auf die Idee kommen* Treffer: 95
- **kommen gar könnte** 33% *könnte ... auf ... Idee kommen ... gar* Treffer: 6
- **kommen gar** 72% *gar nicht erst |auf die Idee [...] kommen* Treffer: 105
- **kommen könnte** 76% *könnte man auf die Idee kommen* Treffer: 366
- **kommen** 88% *auf die Idee [...] kommen* Treffer: 2685

□ **zündende** LLR: 10185 Gesamttreffer: 680

- **zündende hatte gehabt** 100% *Die zündende Idee ... hatte ... gehabt* Treffer: 1
- **zündende hatte** 69% *hatte [... die |eine] zündende Idee* Treffer: 118
- **zündende gehabt** 90% *eine |die zündende [...] Idee [...] gehabt* Treffer: 11
- **zündende hatten** 52% *hatten ... die |eine zündende Idee* Treffer: 21
- **zündende** 88% *eine |die zündende [...] Idee* Treffer: 680

□ **geboren** LLR: 7704 Gesamttreffer: 1627

- **geboren war wurde** 45% *war ... die |als ... Idee [...] geboren [...] wurde* Treffer: 11
- **geboren war Jahren** 66% *vor ... Jahren war die Idee [...] geboren worden* Treffer: 6
- **geboren war** 52% *war [die] Idee [...] geboren* Treffer: 332
- **geboren wurde Jahren** 66% *Vor ... Jahren wurde die Idee [...] geboren ... die* Treffer: 74
- **geboren wurde** 70% *wurde die Idee [...] geboren* Treffer: 760
- **geboren Jahren** 87% *Vor |vor ... Jahren wurde die Idee [...] geboren* Treffer: 105
- **geboren** 84% *wurde die Idee [...] geboren* Treffer: 1627

□ **fixe** LLR: 6964 Gesamttreffer: 597

- **fixe hat** 64% *hat er |eine |die fixe Idee* Treffer: 45
- **fixe ja** 87% *ja ... die |eine fixe Idee* Treffer: 8
- **fixe Nation** 100% *fixe Idee ... einer ... Nation* Treffer: 3
- **fixe** 89% *die |eine fixe [...] Idee* Treffer: 597

□ **Gute** LLR: 2909 Gesamttreffer: 448

- **Gute schlecht** 100% *Gute Idee schlecht* Treffer: 4
- **Gute eigentlich** 100% *Gute Idee eigentlich* Treffer: 11
- **Gute** 96% *Gute [...] Idee* Treffer: 448

□ **absurde** LLR: 2222 Gesamttreffer: 227

- **absurde** sei 84% *Es sei eine absurde Idee sagte ...* Treffer: 19
- **absurde** 96% *auf eine |die absurde [...] Idee* Treffer: 227

□ **pfiffige** LLR: 1803 Gesamttreffer: 170

- **pfiffige hatten** 33% *hatten ... eine pfiffige Idee* Treffer: 6
- **pfiffige ganz** 75% *eine ganz [...] pfiffige Idee* Treffer: 4
- **pfiffige** 89% *eine |die pfiffige [...] Idee* Treffer: 170

□ **Dr** LLR: 1695 Gesamttreffer: 191

- **Dr neueste** 33% *Die neueste Idee ... Dr* Treffer: 3
- **Dr** 46% *die Idee [von ...] Dr* Treffer: 191

□ **neu** LLR: 1421 Gesamttreffer: 950

- **neu nicht ganz** 54% *Die Idee ist nicht ganz neu* Treffer: 53
- **neu nicht** 55% *Die [...] Idee [ist] nicht [...] neu* Treffer: 603
- **neu ganz** 51% *Die Idee ist nicht ganz neu* Treffer: 58
- **neu** 41% *Die [...] Idee [ist nicht] neu* Treffer: 950

□ **olympische** LLR: 1353 Gesamttreffer: 179

- **olympische könnte** 66% *könnte die olympische Idee* Treffer: 3
- **olympische** 79% *die olympische [...] Idee* Treffer: 179

□ **umzusetzen** LLR: 1160 Gesamttreffer: 264

- **umzusetzen Tat auch** 50% *Idee auch in die Tat umzusetzen* Treffer: 4
- **umzusetzen Tat** 98% *die |diese Idee in die Tat umzusetzen* Treffer: 53
- **umzusetzen auch** 70% *die Idee [...] auch [in ...] umzusetzen* Treffer: 24
- **umzusetzen** 95% *die Idee [in die Tat] umzusetzen* Treffer: 264

□ **Regelgestaltung** LLR: 823 Gesamttreffer: 47

- **Regelgestaltung** 100% *Idee 5 |6 Regelgestaltung 6 |5 Ausführung* Treffer: 47

□ **simpel** LLR: 787 Gesamttreffer: 139

- **simpel ganz** 57% *Die [...] Idee ist |war eigentlich ganz simpel* Treffer: 7
- **simpel** 82% *Die [...] Idee [ist] simpel* Treffer: 139

□ **Sozialismus** LLR: 765 Gesamttreffer: 243

- **Sozialismus Freiheit** 75% *Sozialismus in dem |die Idee ... Gleichheit der die ... Freiheit* Treffer: 4
- **Sozialismus schlecht** 90% *den Sozialismus für |sei eine gute Idee die nur schlecht ausgeführt |umgesetzt wurde* Treffer: 10
- **Sozialismus** 67% *der |die Idee [des] Sozialismus* Treffer: 243

□ **ne** LLR: 468 Gesamttreffer: 105

- **ne ja** 80% *ja [mal ...] ne [gute] Idee* Treffer: 10
- **ne** 94% *so ne [...]* Idee Treffer: 105

□ **ja** LLR: 384 Gesamttreffer: 1063

- **ja könnte** 75% *man könnte [...] ja auf die Idee kommen* – Treffer: 48
- **ja** 57% *ja ... eine |auf die Idee* Treffer: 1063

□ **statistisch unspezifisch**

Abb 1: Ausschnitt aus Kookkurrenzprofil von **Idee** (Ausschnitt, zuletzt analysiert am 14.4.2008)¹⁰

Ein solches Kookkurrenzprofil bildet nichts weiter ab die Teilmenge statistisch signifikanter Vorkommen aus einer unspezifischen Vorkommensmenge ab. Statistisch signifikant heißt hier ein auffälliges Kovorkommen zwischen Bezugswort und den Kookkurrenzpartnern in einer bestimmten Umgebung bei einer bestimmten Parameter-Einstellung. Darüber hinaus sind oft rekurrente syntagmatische Strukturen zu erkennen, die nicht beliebig transformierbar sind.¹¹

2.3 Linguistische Systematisierung

Eine linguistische – also qualitative – Verarbeitung der Daten bedeutet nun, inhaltliche Zusammenhänge, Gebrauchsähnlichkeiten, über- und untergeordnete Muster zwischen den berechneten Kovorkommen zu rekonstruieren, diese Beobachtungen zu systematisieren und entsprechende Verallgemeinerungen zu treffen. Diese Systematik soll helfen, die konventionalisierten Gebrauchskonstellationen fester Wortverbindungen aufzudecken und mit

¹⁰ Zur Interpretation der statistischen Angaben (z.B. LLR) verweise ich auf das entsprechende Tutorial des Projektes »Methoden der Korpusanalyse und –erschließung« unter [<http://www.ids-mannheim.de/kl/misc/tutorial.html>] (zuletzt besucht am 29.4.2008).

¹¹ Man kann solche Cluster natürlich auch fokussiert auf die Einzelwortperspektive auswerten. D.h. man interessiert sich primär für Evidenzen der konzeptuellen Relationen von Bezugswort und ihren Kookkurrenzpartnern, um das semantische Potenzial des Bezugsworts z.B. im Sinne von Frames oder thematisch-diskursiven Zusammenhängen zu erfassen (vgl. Firth 1957, Steyer 2004b, Blumenthal 2006). Dieser Zugang fließt beispielsweise in die Bedeutungsbeschreibungen des *lexiko*-Wörterbuch ein [www.owid.de] (zuletzt besucht am 29.4.2008). Die »einzelwortbezogene« Perspektive ist jedoch nicht Gegenstand meiner weiteren Argumentation.

einem unterschiedlichen Grad an Explizitheit zu beschreiben. Jedes vorgefundene Cluster hat einen Aussagewert, den es zu interpretieren gilt.

Demzufolge ist bei diesem Analyseschritt eine spätere Nutzungssituation, die die Bewertung der vorgefundenen Ergebnisse im Vorhinein beeinflussen könnte, zunächst auszublenden. Bei der linguistischen Auswertung von Kookkurrenzprofilen geht es vorrangig um folgende Fragen:¹²

- Welche Binnenstruktur weist ein Kookkurrenzcluster auf?
- Welche typischen Variationen und Gebrauchsrestriktionen lassen sich erkennen (u.a. Festigkeitsgrad/variable lexikalische Ausfüllungen)?
- Gibt es andere Kookkurrenzcluster, die vergleichbare Merkmale und Gebrauchseigenschaften aufweisen?

Abbildung 2 auf S. 128 verdeutlicht, dass verschiedene rekurrente syntagmatische Muster, die im statistischen Kookkurrenzprofil getrennt voneinander und an sehr unterschiedlichen Stellen der statistischen Rangfolge aufgeführt werden, durchaus gemeinsame Merkmale besitzen. Sie bilden über das einzelne Cluster hinausgehende übergreifende Zusammenhänge. Einige Merkmale dieser zusammengefassten Cluster-Gruppen sind morphosyntaktischer Natur wie [PRÄPOSITIONALER_ANSCHLUSS] oder [FLEXIONSFORMEN_{kommen}] und [VERBAL] und teils mit weiteren Merkmalen angereichert wie [ADJEKTIVISCH_{konnotativ}]. Andere dagegen sind eher semantisch begründet wie [VISION_VORSTELLUNG] oder funktional motiviert wie [NEGATION_ZURÜCKWEISUNG]. Das Label [UNSPECIFISCH] fasst alle Vorkommen in dieser Gruppe zusammen, die nicht eindeutig einzuordnen waren und trägt den Charakter eines internen Arbeitskommentars. Natürlich fällt die Heterogenität der Gruppen in Abbildung 3 auf S. 129 sofort ins Auge.

Eine datennahe Interpretation von Kookkurrenzen kann jedoch nicht mit einem einzigen Inventar an beschreibenden Attributen und Kategorien auskommen, z.B. nur ein semantisches Modell oder nur pragmatische Kategorien zu Grunde legen. Kookkurrenzprofile spiegeln die Vielfalt der Sprache wider. Unsere Ordnungskriterien stellen immer eine Kombination aus neuem und tradiertem Wissen, neuen und bekannten Kategorien und Klassen, dar. Ungewöhnliche, bisher noch nicht beschriebene und demzufolge auch nicht klassifizierte sprachliche Erscheinungen kann man zunächst oft nur mit heuristischen Kommentaren fassen, die in ihrer Vagheit noch keinen kategorialen Status haben. In vielen Fällen lassen sich die beobachteten Phänomene durchaus gut mit bereits eingeführten Kategorien fassen. Insofern relativiert sich hier das zu Beginn der Argumentation negierte A-Priori-Prinzip in der qualitativen Analysephase in gewisser Weise natürlich.

Wie eine interne Differenzierung einer einzelnen Syntagmengruppe aussehen kann, illustriert die Abbildung 4 auf S. 130. Die Abbildung stellt nun bereits eine Verallgemeinerung zu Wortverbindungen dar, die auf der Basis der systematisierten KWICs vorgenommen wurde. Aus phraseologischer Perspektive interessieren uns im Folgenden genau solche Einheiten wie in der Abbildung 3, bei denen eine über das reine lexikalische Kovorkommen hinaus gehende **rekurrente syntagmatische Qualität** zu erkennen ist. Rekurrente syntagmatische Qualität bedeutet, dass nicht nur bestimmte lexikalische Einheiten häufig in einer bestimmten Spanne miteinander auftreten, sondern dass dieses Vorkommen immer auch

¹² Eine vergleichbare methodische Herangehensweise verfolgt Duffner (2006), der Satzadverbien auf der Basis von Kookkurrenzprofilen analysiert.

mit einem Mindestmaß an typischer, d.h. wiederkehrender syntaktischer Bindung einhergehen muss. Das Konzept der »usuellen Wortverbindungen« bietet hierfür den Erklärungsrahmen (vgl. Steyer 2000 und 2008).¹³

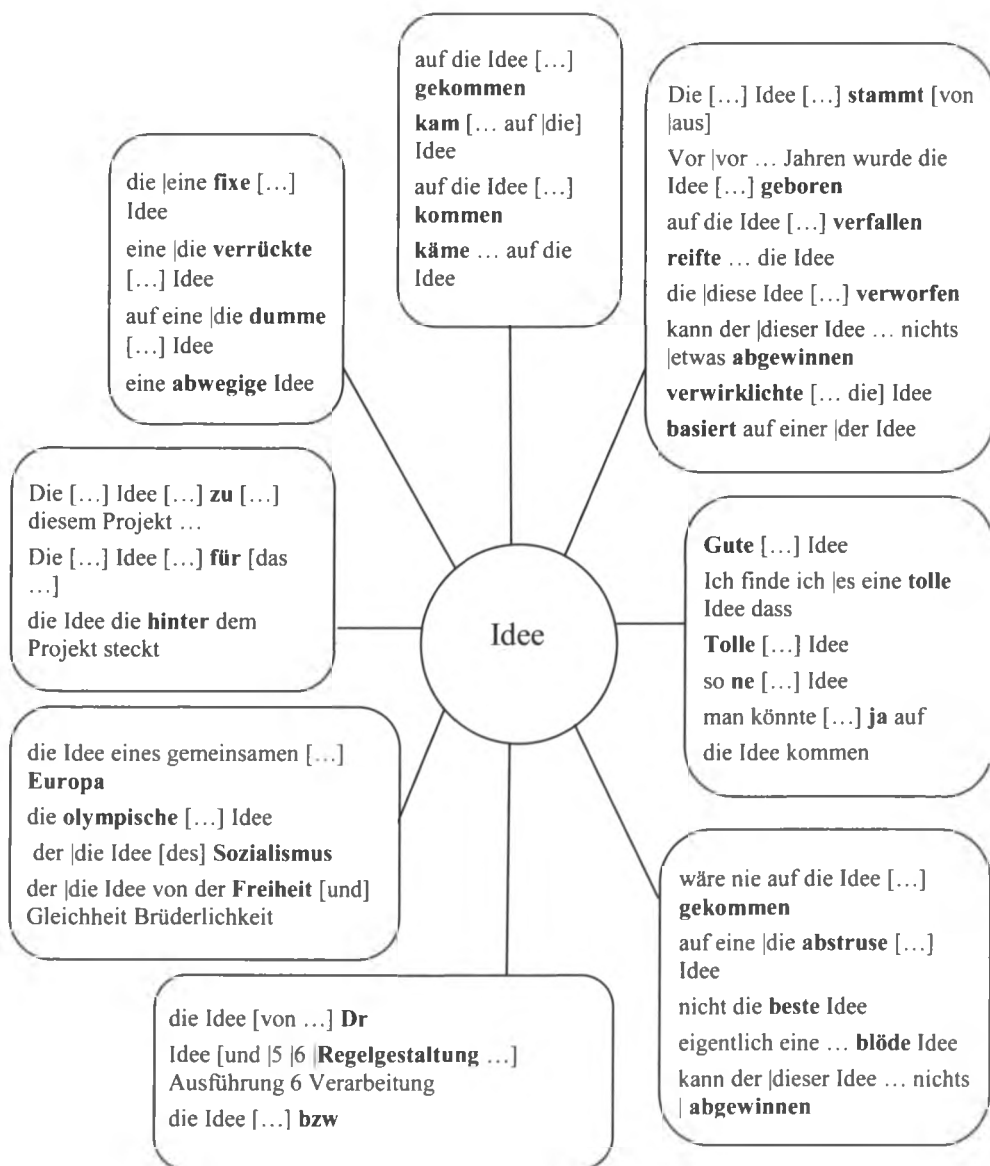


Abb. 2: Systematisierung rekurrenter syntagmatischer Muster

¹³ In Steyer (2008) wird der UZW-Ansatz komplementär zu den Ausführungen in diesem Band aus einem anderen Blickwinkel diskutiert, nämlich der Kollokationsperspektive im Sinne des Hausmannsches Modells.

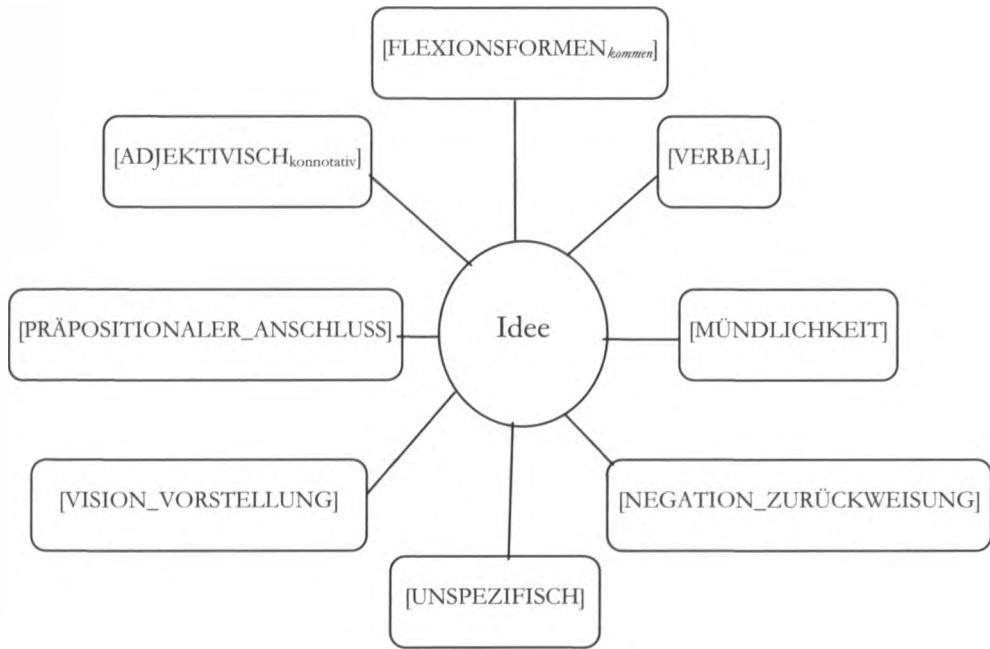


Abb. 3: Systematisierung der den jeweiligen Knoten zu Grunde liegenden Merkmale

3. Zur Bestimmung usueller Wortverbindungen

Usuelle Wortverbindungen (UWV) sind konventionalisierte Muster des Sprachgebrauchs, die in rekurrenten, also wiederkehrenden, syntagmatischen Strukturen manifest werden. UWV haben einen zumindest binären Einheitsstatus und sind minimal lexikalisch ausgefüllt. Die Komponenten einer usuellen Wortverbindung können unterschiedliche Ausformungen haben, von der einzelnen Wortform eines sprachlichen Ausdrucks bis zu Ausdrucksklassen. Ihre Übersummativität ist nicht eng semantisch zu interpretieren, etwa im Sinne der Dichotomie *kompositionell* vs. *nicht-kompositionell* oder *idiomatisch* vs. *nicht-idiomatisch*, sondern primär pragmatisch. In diese Definition habe ich zwei Bestimmungsmerkmale für Phraseologismen, die die strukturelle Fixiertheit betreffen, integriert: Polylexikalität und Festigkeit (vgl. Burger 2007). Idiomatizität wird dagegen als ungeeignet für die Festlegung des Objektbereichs angesehen, da sie – wie gezeigt – ein relationales Kriterium ist, das auch nur eine sprachliche Ebene, nämlich die eng semantische, fokussiert.

Zugleich wird Feilkes Konzept der idiomatischen Prägung (u. a. 1996, 2004) als zentrale Instanz aufgenommen. »Idiomatisch« ist in diesem Kontext nicht als kognitiv-semantische Qualität zu verstehen, wie sie vor allem Dobrovol'skij mit seinem Idiom-Konzept vertritt (vgl. 1995), sondern im Sinne einer konventionalisierten Gebrauchsnorm (vgl. auch Stubbs 1997). UWV sind somit »funktionale Chunks«. UWV werden nun unter den beiden zentralen Bestimmungsmerkmalen »lexikalische Spezifiziertheit« und »funktionale Gebundenheit« weiter betrachtet.

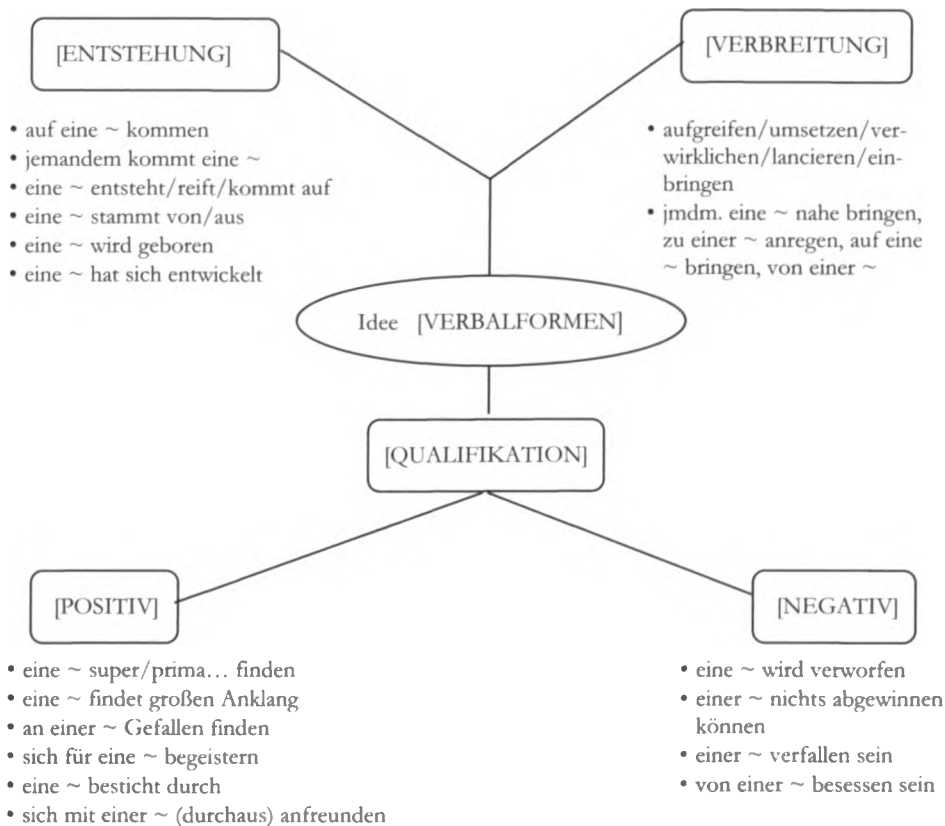


Abb. 4: Inhaltliche Gruppierung von Syntagmen

3.1 Lexikalische Spezifiziertheit

Ein geeignetes Kriterium zur Klassifikation von usuellen Wortverbindungen auf der Basis von Kookkurrenzprofilen ist der jeweilige Grad an lexikalischer Spezifiziertheit des vorgefundenen syntagmatischen Musters. Es ist zu fragen, inwieweit ein statistisch signifikantes Syntagma unikal lexikalisiert ist, in Hinblick auf seine lexikalischen Komponenten skalar variabel oder ob es nur eine – wenn auch typische – Realisierung anderer Bildungsmuster in ein- und demselben Paradigma repräsentiert.¹⁴ Durch die Kookkurrenzanalyse kann man Hinweise auf unikal lexikalisierte Syntagmen erhalten.

¹⁴ Deppermann unterscheidet in Bezug auf den »Grad idiomatischer Formfixiertheit« ebenso zwischen lexikalisch voll spezifizierten, lexikalisch teilspezifizierten und voll schematisierten Konstruktionen (2006: 48f.) Für UWV sind die ersten beiden Typen relevant. Legt man das Begriffsspektrum der Konstruktionsgrammatik – so wie von Fischer/Stefanowitsch (2006b) beschrieben – zu Grunde, befindet sich das UWV-Konzept auf einer mittleren Ebene. Das UWV-Konzept schließt zum einen rein wortbildungsbezogene Konstruktionen und Einwortlexeme aus, zum anderen »abstrakte, lexikalisch nicht ausgefüllte syntaktische Strukturen, wie phrasale Kategorien,

Diese UWV treten in all ihren Korpusvorkommen als lexikalisch und syntaktisch fixierte Phrasen hervor. Das bedeutet, dass im Prinzip nur Muster mit denselben lexikalischen Vertretern identifiziert wurden. In den seltensten Fällen handelt es sich um eine absolute Fixierung in dem Sinne, dass nur ein Exemplar einer Zeichenkette mit genau einem anderen Exemplar einer Zeichenkette in genau einer syntaktischen Struktur korreliert und mit einer und derselben Bedeutung und Funktion. Zumeist gibt es auch bei diesem hohen Grad an Lexikalisierung minimale Varianzen in Schreibung, Stellung oder Flexion. Diese Varianz kann den autonomen Status dieser Wortverbindung unangetastet lassen, kann aber auch zu Differenzierungen im Gebrauch führen. Beispiel für diese Gruppe der unikal lexikalisierten usuellen Wortverbindungen sind in (2) und (3) aufgeführt:

- (2) *fixe Idee* (siehe Punkt 2.1.)
nichts ist stärker, als eine Idee, deren Zeit gekommen ist (satzwertiges Zitat)
- (3) *eine Idee* (in der Bedeutung eines mehrgliedrigen Quantifikators)
Keine Idee? (in der satzwertigen Ellipsenform)

Diese lexikalische Fixiertheit kann sich sowohl auf Relationen zwischen autosemantischen Komponenten oder auf mehrgliedrige (auch satzwertige) Phrasen wie in (6) als auch auf syntagmatische Minimaleinheiten wie in (7) beziehen. Syntagmatische Minimaleinheiten sind statistisch signifikante Verbindungen zwischen einem Inhaltswort und einem/mehreren Funktionswörtern. Das Kriterium für den UWV-Status solcher Minimaleinheiten ist ihre Autonomie als sprachlicher Ausdruck. Das kann eine eigenständige Bedeutung im engen Sinne als auch auf eine eigenständige Funktion in der Kommunikation bedeuten. So lässt sich die Wortverbindung *eine Idee* – neben der unspezifischen Wortkombination [ARTIKEL_{UNBESTIMMT}]+*Idee* – mit *nur ein wenig, ein kleines bisschen, ein Quäntchen* usw. paraphrasieren. Typische Muster sind z.B. *eine Idee braver, ernsthafter, raffinierter, schneller; mehr X* (z. B. *groovt und knallt es, Profil, Zuneigung*). Für die satzwertige Ellipse *Keine Idee?* könnte man alternativ andere Formulierungen verwenden, die jemandem zum Einbringen einer Idee auffordern/animieren wie *Und fällt dir dazu nichts ein? Hast Du einen Vorschlag?* usw. Diese Autonomie lässt sich also mittels Paraphrasen oder Ersetzungen verifizieren. Die orthografische Variante *keine Idee* ist dagegen in der Regel in Verbalphrasen eingebettet (*keine Idee haben, keine gute/schlechte Idee sein*) oder auch als elliptisches Element *keine Idee, warum...* Jede dieser Varianten unterscheidet sich hinsichtlich Bedeutung und/oder illokutivem Potenzial von der elliptischen Verwendungsweise mit Satzwertigkeit: *keine Idee haben* ‘keinen passenden Gedanken zu etwas haben’, *keine gute Idee sein* ‘den Gedanken eines anderen zurückweisen’; *keine schlechte Idee sein* ‘den Gedanken/Vorschlag eines anderen in die eigenen Überlegungen einbeziehen, wohlwollend betrachten’; *keine Idee, warum...* ‘den Gedanken/Vorschlag eines anderen nicht verstehen, nicht akzeptieren (wollen)’.

Die Kookkurrenzanalyse erbringt vor allem aber auch bereits Evidenzen zu lexikalisch teilspezifizierten Wortverbindungen. Dieser Typ ist der Normalfall und stellt eine Abstraktion über konkrete lexikalische Realisierungen dar. D.h. die Fixiertheit lässt sich je nach Korpusbefunden mit einem unterschiedlichen Grad an Verallgemeinerung erklären.

Argumentstrukturen (...).« (Fischer/Stefanowitsch 2006b: 6). Die prinzipielle Klärung des Verhältnisses des UWV-Ansatzes und den einschlägigen Konzepten der *Construction Grammar* muss anderen Publikationen vorbehalten bleiben.

Die Flexibilität bezieht sich auf die Austauschbarkeit lexikalischer Vertreter ähnlicher Art in einem Paradigma. Die Ausdehnung kann von der minimalen Ersetzbarkeit eines anderen lexikalischen Vertreters ähnlicher Art bis hin zur Fixiertheit auf der abstrakten Ebene von Ausdrucksklassen gehen. Auf diese Weise lassen sich graduell lexikalisierte Wortverbindungen unterscheiden:

- (a) **Vorkommen mit minimaler lexikalischer Varianz:** Zu dieser Gruppe zählen Wortverbindungen, die nur mit einigen wenigen lexikalischen Realisierungsformen signifikant vorkommen. So ist z.B. *auf die/eine Idee kommen* hochgradig fest. Analysiert man das Kookkurrenzverhalten des Musters *auf X kommen*, ist nur das Nomen *Gedanke* in derselben Bedeutung 'etwas einfallen, getan haben' auffällig häufig (*auf einen Gedanken kommen*). Die anderen nominalen X-Füller indizieren andere Muster mit anderen Bedeutungen (z.B. *auf Spur/Schliche kommen* = 'aufdecken'; *auf Touren kommen* 'sich endlich entwickeln, bewegen; (endlich) einen Fortschritt erreichen' oder *auf den Tisch kommen* = 'Probleme endlich zur Sprache bringen'). Die Vorkommen dieser Gruppe sind noch im Übergangsbereich zu unikal lexikalisierten Wortverbindungen anzusiedeln.
- (b) **Vorkommen mit eingeschränkter lexikalischer Varianz:** Bei diesem Varianztyp sind mehrere lexikalische Füller zu identifizieren, die auf eine potenzielle Bildbarkeit mit einer spezifischen Ausdrucksklasse hindeuten. Einer dieser lexikalischen Vertreter ist jedoch auffallend signifikant. So wird die Wortverbindung *eine Idee ist/wird/war geboren* mit Abstand am häufigsten realisiert. Andere Nomen, die dieses Muster auch ausfüllen, sind aber im Verhältnis zum Lexem *Idee* viel seltener im Korpusgebrauch: *X wird geboren* → *X = Plan, Einfall, Gedanke, Vorhaben, Absicht* usw.
- (c) **Vorkommen mit systematischer lexikalischer Varianz:**

Diese Wortbindungsmuster lassen sich als echte Bildungsmuster charakterisieren, die eine systematische Varianz von vielen lexikalischen Vertretern ähnlicher Art innerhalb eines Musters aufweisen:

- (4) [ADJEKTIV_{positiv}]¹⁵ + *Idee / Ideen*
Typische lexikalische Vertreter = *gut/zündend/toll/schön/großartig/genial/pfiffig/brillant/faszinierend/bahnbrechend/bestechend/glänzend*
- (5) [QUANTIFIKATOREN] + *Ideen*
Typische lexikalische Vertreter = *eine Fülle von ~/eine Menge ~ /eine Vielzahl von ~ /viele ~ /ein paar ~ /voller ~*
- (6) [EINHEITEN, DIE EINE ZURÜCKWEISUNG AUSDRÜCKEN] + *Idee*
Typische lexikalische Vertreter = *keine wirklich gute ~/eine absurde, abstruse, abwegige ~/auf eine solche ~ muss man erstmal kommen/niemand wäre auf eine solche ~ gekommen/eigentlich eine blöde ~*

¹⁵ Es handelt sich hier um Adjektive, die eine positive Konnotation in sich tragen.

Ein großes Problem bei der Festschreibung eines Wortverbindungsstatus ist die empirische Beobachtung, dass sich in den meisten Wortverbindungen mehrere Bildungsmuster überlappen. Mit korpusgesteuerten Analysen stoßen wir auf ein kompliziertes System »verschachtelter« Wortverbindungen, das von einer Vielzahl wechselseitiger Vernetzungen und Hierarchien geprägt ist. Auch der Status einer Einheit als Ausgangswort relativiert sich unter diesem Blickwinkel. Werden die entdeckten Muster ebenfalls einer automatischen Umgebungsanalyse unterzogen, zeigt sich oft, dass unser eigentliches Bezugswort auch nur ein typischer Vertreter ähnlicher Art, ein typischer Füller ist. Dies wird in den Beispielen 4a, 5a, 6a dargestellt:

(4a)

Idee-Wortverbindung: *brillante Idee*
 Externes Muster: **brillant X**
 X-Füller: *Rhetorik/Technik/Analyse/Dialoge/Idee/Inszenierung/Köpfe/*
 Ausdrucksklasse: [NOMINA_{UNBELEBT}, DIE KREATIVITÄT ODER KREATIVE ERGEBNISSE
 AUSDRÜCKEN]

(5a)

Idee-Wortverbindung: *eine Fülle von Ideen*
 Externes Muster: **eine Fülle von X**
 X-Füller: *Informationen/Möglichkeiten/Daten/Aufgaben/Ideen/Angeboten/
 Chancen/Problemen/Tipps/Eindrücken/Vorschriften*
 Ausdrucksklasse: [NOMINA_{QUANTIFIZIERBAR}]

Idee-Wortverbindung: *voller Ideen*
 Externes Muster: **voller X**
 X-Füller: *Tatendrang/Elan/Ideen/Eifer/Energie*
 Ausdrucksklasse: [NOMINA, DIE AUSDRUCK AUSGEPRÄGTEN HANDELNS UND WILLENS SIND]

(6a)

Idee-Wortverbindung: *eigentlich eine blöde Idee*
 Externes Muster: **eigentlich ein/eine** [NEGATIVE KONNOTATION X]¹⁶
 X-Füller: *Unding/alter Hut/Fehlbesetzung/Witz/Widerspruch in sich/Schande/
 eine blöde Idee/Skandal/ein Ding der Unmöglichkeit*
 Ausdrucksklasse: [AUSDRÜCKE, DIE EINE NEGATIVE KONNOTATION TRANSPORTIEREN UND
 EINE ZURÜCKWEISUNGSHANDLUNG INDIZIEREN]

Völlig offen ist noch, auf welcher Abstraktionsebene sich unsere kognitiven Verarbeitungen und Aktualisierungen nun wirklich vollziehen und welche Muster wann mit welchen lexikalischen Bezugseinheiten aktualisiert werden. Sicherlich sind diese Verarbeitungen stark kontextabhängig. Möglicherweise werden je nach Kommunikationssituation bestimmte sprachliche Einheiten zu konzeptuellen Bezugswörtern, die dann auch das lexikalische Zentrum der Wortverbindung bestimmen. Die Ermittlung des invarianten Status vs. varianter Elemente und der typischen Realisierungen von Bildungsmustern verschiedener

¹⁶ Natürlich lässt sich auch das positiv konnotierte Muster ansetzen: *eigentlich ein/eine* [POSITIVE KONNOTATION X]; X-Füller z.B. *ein ganz normales Leben/ein guter Kerl/eine gute Idee/Sache; ein gutes Zeichen*.

Abstraktionsstufen kann bis zu einem gewissen Maße auf der Basis automatischer Analysen und Clusteringverfahren vorgenommen werden. Die Qualifizierung der Typen von Ausdrucksklassen, die jeweils miteinander korrelieren, der Eigenschaften struktureller Restriktionen und funktionaler Gebundenheiten bleibt nach wie vor bisher allein dem interpretierenden Linguisten vorbehalten. Aber anders als früher kann er diese Interpretationsleistung auf der gesicherten analytischen Basis sprachlicher Massendaten erbringen.

3.2 Funktionale Gebundenheit

UWV sind als holistische **kommunikative Entitäten** zu interpretieren, die als zentrale Konstituenten des Sprachgebrauchs fungieren (vgl. dazu »semantisch-pragmatische Einheiten« bei Siepmann 2002). Tomasello (2006), der Konstruktionen im frühen Erstspracherwerb untersucht, beschreibt die »funktionale Verfestigung« von Wortclustern durch den massenhaften Gebrauch so (zit. nach Übersetzung von Wulff/ Zeschel in Fischer/ Stefanowitsch 2006a: 21):

Aus der gebrauchsgestützten (*usage-based*) Perspektive der Konstruktionsgrammatik ist Sprache zunächst im Kontext der grundlegenden Tatsache zu betrachten, dass sich Menschen zu bestimmten Anlässen mit bestimmten, aus konkreten sprachlichen Bausteinen bestehenden Äußerungen verständigen. Wenn Menschen wiederholt »ähnliche« Dinge in »ähnlichen« Situationen sagen, entwickelt sich daraus mit der Zeit ein sprachliches Verwendungsmuster, das in den Köpfen der Benutzer als neue Kategorie oder Konstruktion schematisiert wird – mit unterschiedlichen Abstraktionsgraden. [...] Es gibt keine sprachlichen Elemente – ob lexikalische oder syntaktische, konkrete oder abstrakte, reguläre oder idiomatische – die nicht in diesem Sinne symbolisch sind; alle haben einen kommunikativen Bedeutungsgehalt, weil sie alle direkt aus dem Sprachgebrauch abgeleitet sind.

Feilke (2004: jeweils 47 u. 43) formuliert eine pragmatisch motivierte Verfestigung u.a. auf folgende Weise:

Wortverbindungen werden fest durch den Gebrauch und sind pragmatisch fixiert innerhalb konventionaler Gebrauchskonstellationen.

Pragmatisierung soll heißen, dass das Spektrum idiomatischer Prägung extensional nicht mehr über strukturlinguistisch zu fassende Irregularitäten, sondern zunehmend über das Kriterium der pragmatisch usuellen semiotischen Einheiten des Sprachgebrauchs definiert wird.

Im Mittelpunkt stehen die Beschränkungen, denen diese Wortverbindungen im Gebrauch unterliegen. Diese Gebrauchsrestriktionen können sich, so Deppermann (2006: 48), sowohl auf die »semantischen Klassen möglicher lexikalischer Instanzierungen« beziehen als auch auf die speziellen, eingeschränkten »syntaktischen oder pragmatischen Kontextbedingungen«. Die folgenden Analysebeispiele sollen illustrieren, wie sich Kookkurrenzrelationen zwischen lexikalischen Einheiten unter funktionalen Gesichtspunkten und unter Einbeziehung rekurrenter Kontexte ausdifferenzieren. Zugleich soll deutlich werden, inwieweit funktionale Muster durch ganz unterschiedliche Versprachlichungen realisiert werden können. Eine klare Trennung zwischen Bedeutung und Funktion ist bei einer solchen Herangehensweise nahezu nicht mehr möglich. Daher verwende ich im Folgenden die heuristische Bezeichnung »Funktionsbedeutung« und unterscheide nicht zwischen der Bedeutung in einem eng semantischen Sinne und der realisierten kommunikativen Funktion.

Die durch die automatische Kookkurrenzanalyse berechneten Primär-Cluster zum Bezugswort *Idee* mit den primären Kookkurrenzpartnern *gekommen, kommen, kam, käme* (vgl. Abbildung 1 in Punkt 2.2.) liefern uns zwar Indizien für lexikalische Affinitäten, sind aber in Bezug auf die pragmatischen Kontexte ambig. Das bedeutet, dass die primäre lexikalische Korrelation zwischen zwei lexikalischen Einheiten in der Regel noch nichts über die eigentlich realisierten kommunikativen Funktionen aussagt. Durch die Subcluster und die syntagmatischen Muster erhält man dann erste Evidenzen. Den eigentlichen Erkenntniswert liefern aber die dahinter liegenden Originaltextausschnitte (KWICs, Volltextstellen), hier die KWICs des Clusters *Idee – gekommen* als Minimal-Ausschnitt aus der natürlichsprachigen syntaktischen Einbettung:

- HMP06 Um es nicht so streng wirken zu lassen, war jemand auf die **Idee** gekommen, rote, blaue, gelbe und grüne Neonröhren an den Fassaden
- HMP06 Sollte Kaul auf die **Idee** gekommen sein, sie überwachen zu lassen?
- HMP06 Wie sind Sie auf die **Idee** gekommen, sich eine blutige Nase zu verpassen? Reinking: Ich habe
- HMP06 seinen Kindern im Kino und wundert sich: »Ich bin nicht auf die **Idee** gekommen, dass die Kinos so unverschämt sind, die Kinder so«.
- REI Sie erreicht!) </footnote> Sie sind nicht ein einziges Mal auf die **Idee** gekommen, dass es bei einem Anteil von nur noch rund 23 Prozent
- REI Sie dem Hohen Hause wirklich erklären, wie Sie jetzt auf diese **Idee** gekommen sind. Es gibt intelligentere Lösungen, die mehr auf den
- REI diese Glocken jetzt zehn Monate lang geschrillt, ehe Sie auf die **Idee** mit dem Untersuchungsausschuss gekommen sind. <footnote>(Beifall
- REI Plakate geklebt, auf denen stand: »Nichts ist mächtiger als eine **Idee**, deren Zeit gekommen ist.«
- SPK obwohl er zugibt, dass er »niemals selbst auf die **Idee** von Bakterien in Schlüpfern gekommen wäre«.
- SPK Wären die Zellen nicht auf die grandiose **Idee** dieser Verpackung gekommen, sie hätten ein Problem.
- SPK »begegnet wird«, erklärt Page. Wann genau das Y-Chromosom auf die **Idee** der Selbstkontrolle gekommen ist, können die Wissenschaftler noch
- SPK Hobbies seien Lesen und das Surfen im Internet, wobei ihr auch die **Idee** zu ihrem Vorschlag gekommen sei. Serena darf sich nun auf einen
- SPK viele Gruppen unserer Vorfahren unabhängig voneinander auf die **Idee** gekommen, die schmackhaften und genügsamen Borstenviecher zu
- SPK Nichts ist eben mächtiger, wusste schon Victor Hugo, »als eine **Idee**, deren Zeit gekommen ist«. Andreas Jahn <url
- SPK schneller auswerten zu können, waren Forscher damals auf die **Idee** gekommen, diese Aufgabe auf viele Rechner zu verteilen. Dafür
- FSP Goethe, Schiller, Heine usw. wird damals wohl kaum jemand auf die **Idee** gekommen sein, Männern generell gewisse sprachliche Fähigkeiten
- VDI06 man Gefühle in die Chemie einbringen kann. So bin ich auf die **Idee** gekommen, für mein Unternehmen einen Workshop mit Künstlern und
- VDI06 den Möglichkeiten des Werkstoffs Kupfer wären wir nicht auf die **Idee** gekommen, die CIS-Solarzelle (Kupfer/Indium/Selen) bis zur
- VDI06 Maschinenbaus zu sprechen. VDI nachrichten: Wie sind Sie auf die **Idee** gekommen, solch ein Buch zu machen? Salié: Der Erfolg unseres

Abb. 5: Rekurrente syntagmatische Qualität

Betrachtet man die KWICs, wird deutlich, dass sogar innerhalb ein- und desselben Kotextclusters – trotz identischer lexikalischer Affinität – sehr verschiedene Funktionsbedeutungen zu lokalisieren sind. Diese Funktionsbedeutungen findet man dann auch in den anderen Clustern wieder: Für die Cluster *Idee* – Flexionsformen des Lemmas *kommen* sind es z. B. folgende Typen:

(7)

Typ: ENTSTEHUNG

- T92 Spaziergängen durch den Englischen Garten **kam** Rainer auf die **Idee**, daß ich eine Schauspielagentur eröffnen soll.
 E97 sind wir Grünen auf die **Idee** des »Misstrauens auf Halde« **gekommen**.
 T90 antiideologisch und übernational. »Da kam mir spontan die **Idee**, so etwas müßte doch auch in Zürich möglich oder sogar nötig sein.
 SPK schneller auswerten zu können, waren Forscher damals auf die **Idee gekommen**, diese Aufgabe auf viele Rechner zu verteilen.

Beispiel für *Idee*-Wortverbindungen: *jemandem ist die Idee gekommen, jemand ist auf eine [KONKRET BENANNT] Idee gekommen*

Funktionsbedeutung: Sprecher beziehen sich auf die Situation, in denen ein Gedanke/Einfall entstanden ist (oft plötzlich oder unerwartet,) oder auf die jeweils beteiligten Personen (selbst- oder fremdreferenziell).

(8)

Typ: ZURÜCKWEISUNG

- O94 stolz. Einer Hausfrau wäre diese **Idee** wohl nie **gekommen**.
 T92 Niemand wäre auf die **Idee gekommen**, in Uganda einzugreifen.
 N99 zu haben. Auf so eine **Idee** muss man erst **kommen**. Klima
 M98 »Kein Mensch wäre auf die **Idee** gekommen, die Frage, Gibt es
 T05 **käme** kein Zuschauer auf die **Idee**, bereits vorher zu klatschen.

Beispiele für *Idee*-Wortverbindungen: *wäre/wären nie auf die Idee gekommen; würde erst gar nicht auf die Idee kommen; wohl niemand wäre auf diese Idee gekommen/auf die Idee muss man erstmal kommen*

Funktionsbedeutung: Sprecher drücken damit die Absurdität eines Gedankens als Zurückweisung aus, oft verbunden mit einer Kritik an einer Person.

(9)

Typ: OPTION

- P99 könnte ein Gericht auf die **Idee kommen**, diese Produktgruppe auszugliedern.
 B01 aussprechen, könnte jemand auf die **Idee kommen**, uns die Kosten aufzuerlegen,
 HMP06 besser? Keine Frage, auf die **Idee** könnte man **kommen**. Aber

Beispiele für *Idee*-Wortverbindung: *könnte man auf die Idee kommen*

Funktionsbedeutung: Sprecher formulieren damit ein Gedankenspiel, bei dem eigentlich absurde oder nicht denkbare Vorstellungen als mögliche Optionen thematisiert werden. Häufig fungieren solche Wortverbindungen auch als argumentative Figur des Abwägens von Argumenten.

(10)

Typ: FRAGE/ZWEIFEL

- T89 wie seid ihr auf die **Idee gekommen**? Frank: Wir selber haben
 F03 Und deshalb sind Sie auf die **Idee gekommen**, die bestehenden
 Regeln viel strenger auszulegen? Ja.
 T02 »Wie kann ein Rezzo Schlauch nur auf die **Idee kommen**, diese Affäre ohne Rücktritt
 überstehen zu können?«

Beispiele für Idee-Wortverbindung: *auf die Idee gekommen sein, auf die Idee kommen können*

Funktionsbedeutung: Sprecher fragen nach den Motiven für einen Gedanken oder eine Handlung. Oft drücken sie damit einen zumindest leichten Zweifel an der Sinnhaftigkeit eines Gedankens/Vorschlags bzw. der Handlung aus.

Diese Funktionszuschreibungen lassen sich zumeist – wie an den Beispielen deutlich wird – durch grammatische Marker wie Satzmodus, Negationsstruktur oder Modalität im Kotext verifizieren. Interessant sind nun aus korpusgesteuerter Sicht vor allem die Fälle, in denen eine solche Funktion durch andere oder auch unerwartete sprachliche Realisierungen indiziert wird. So sind die Funktionsbedeutungen (7) – (10) nicht nur in den Clustern *Idee – kommen*_{LFMMA} zu identifizieren, sondern sind auch in anderen sprachlichen Vorkommen im Kookkurrenzprofil von *Idee* zu finden. Das folgende Beispiel zeigt das Spektrum lexikalischer Varianz bei gleichzeitiger funktionaler Invarianz:

(11)

[BESTÄTIGUNG/ZUSTIMMUNG/AKZEPTANZ]

*wirklich eine gute Idee
 wäre gar keine schlechte Idee
 Gute/Prima/Super/Tolle Idee!
 die Idee finde ich super
 halte ich für eine gute Idee
 zündende/tolle/schöne/großartige/geniale/pfiffige/brillante/faszinierende/bahnbrechende/
 bestechende/glänzende Idee
 eine ebenso so einfache/simple wie geniale Idee
 die Idee ist zwar nicht ganz neu, aber
 könnte mich mit der Idee anfreunden
 ist ja mal ne gute Idee
 eine glorreiche/grandiose Idee*

(12)

[ZURÜCKWEISUNG/KRITIK]

*Idee klingt gut/einleuchtend, aber
 keine wirklich gute Idee
 wohl nicht die beste Idee
 eigentlich keine schlechte/eine hübsche/ganz nette Idee, aber*

fixe/wahnwitzige/aberwitzige/abwegige Idee
eine glorreiche/grandiose Idee

Vielen dieser Wortverbindungen liegen abstraktere Bildungsmuster zu Grunde, z.B.:

(13)
wirklich eine gute X sein
klingt X, aber...
wohl nicht X sein
eine ebenso X wie Y sein
ist ja mal ein X
eigentlich kein X sein

Beispiel (14) zeigt, dass viele Funktionsbedeutungen überhaupt erst aus dem weiteren Kontext ersichtlich werden, z. B. bei der Umdeutung konnotativer Adjektive durch Negation (z.B. *eigentlich keine schlechte Idee sein*).

Wie differenziert die Kontextmuster sein können, zeigen die Einbettungen der eigentlich positiv konnotierten Wortverbindungen *glorreiche Idee* in Kontexte, die eine negative Konnotation hinzufügen (analog dazu auch die Kontextmuster für *grandiose Idee*).

Kontextuell bedingte negative Bewertungen treten z.B. durch folgende sprachliche Phänomene hinzu:

(14)
 (a) Argumentationsmuster

<ZWAR> trifft X zu <DENNOCH> niemand g. *Idee*

Heute ist er bei nur noch knapp 2.000 aktiv Beschäftigten und 3.000 Kurzarbeitern zu klein geworden. Und dennoch käme niemand auf die **glorreiche Idee**, ihn zu vergrößern. Denn schon in wenigen Monaten könnte er dicke ausreichen, vielleicht sogar leer und trostlos das Ende verkünden. (*die tageszeitung*, 07.05.1991, S. 13; »Die Lektion von Greifswald«)

(b) Lexikalische Marker

leider

»Wer hatte denn die Idee mit den Pickerln?«, fragt Richterin Brigitte Zeilinger. »Das war **leider** meine **glorreiche Idee**«, meldet sich L. (*Die Presse*, 27.11.1996)

auf~ verfallen

Sie werden auch nicht billig sein. Angesichts der leeren Staatssäckel könnten die Abgeordneten auch auf die **glorreiche Idee verfallen**, nicht das Einkommensteurrecht [sic!] zu ändern, sondern die Verfassung. Dann hätten nicht nur die Verfassungsrichter das Nachsehen, sondern vor allem die Familien. (*Vorarlberger Nachrichten*, 22.10.1997, S. D6)

Anführungszeichen

Die Hausbesitzerin, eine ältere Dame, wollten die Burschen gegen 4.30 Uhr auch nicht herausläuten. So kamen sie laut Gendarmerie auf die »**glorreiche Idee**«, den Freund über das Fenster abzuseilen. Als der Student durch das Fenster in eine Schlinge des Seils stieg, riß es. (*Salzburger Nachrichten*, 03.05.1997)

(c) Thematisierung von Ursache-Wirkung/Folge

Die nach dem Mauerfall in übertriebener Eile geschlossene künstlerische Ehe mit dem Chefdirigenten Michael Schoenwandt zeigte unübersehbar Ermüdungserscheinungen, einige Solo-Stellen schienen kaum mehr adäquat besetzt. **Ausgerechnet in dieser angepannten** [sic!] **Situation** hatte der seinerzeit amtierende Kultursenator Peter Radunski die **glorreiche Idee**, das symphonische Ost-Pendant zum Berliner Philharmonischen Orchester mit dem Orchester der Komischen Oper zu verschmelzen. Dem BSO drohte **nunmehr gar die Auflösung**, die Stimmung sank auf den Nullpunkt. (*Frankfurter Allgemeine*, 27.03.2001)

(d) Rhetorische Fragen/Ironie

Eine Sendung, in der Waldemar Hartmann Olympioniken interviewt und Beiträge über ihren Tag ammoderiert, ist ja an sich nicht abwegig. **Aber was macht eigentlich derjenige beruflich**, der auf die **glorreiche Idee** gekommen ist, dabei Harald Schmidt neben die gefürchtete fränkische Duzmaschine zu setzen? (*Mannheimer Morgen*, 14.02.2006)

Die Bilder, die die Erwachsenen auf den Handys sehen konnten, trieben ihnen die Schamesröte ins Gesicht. Um das Problem zu lösen, kamen Bildungspolitiker auf die **glorreiche Idee** Mobiltelefone auf dem Schulhof zu verbieten.

Ein toller Einfall. (*Berliner Zeitung*, 03.04.2006, S. 26)

Es sei noch einmal festgehalten: Typen und Realisierungen von Funktionen in der Kommunikation sind nicht abhängig vom Grad der Lexikalisierung einer Wortverbindung.

Sie können sowohl unikalene Wortverbindungen als auch – als Schemafunktion – ganzen Wortverbindungsmustern zugeschrieben werden. Sprecher aktualisieren jeweils die passenden sprachlichen Bausteine. Sie adaptieren die unterschiedlichen Komponenten ein und desselben pragmatischen Konzepts situationsabhängig.

Inwieweit die in diesem Kapitel entwickelten Erklärungsansätze für UWW für praxisnahe Anwendungen wie die Mehrwortlexikografie fruchtbar gemacht werden können, wird sich in der Zukunft zeigen. In jedem Falle liefern sie wichtige Evidenzen zur Typik des Gebrauchs von Wortverbindungen, die in eine lexikografische Beschreibung unbedingt einzufließen haben (vgl. dazu noch einmal Punkt 5).

Dass sich auf der Basis solcher datennaher methodischer Vorgehensweisen auch neue Perspektiven für Darstellungs- und Präsentationsformen ergeben, soll im nächsten Kapitel zumindest kurz skizziert werden.

4. Formen der Beschreibung

Durch das korpusgesteuerte Vorgehen rücken die realen Sprachdaten nicht nur in neuer quantitativer Dimension in den Fokus der Analyse, sondern sie können selbst zu Hauptkomponenten der Präsentation der Ergebnisse werden. Eine intelligente Systematisierung der automatisch ermittelten signifikanten Kontextmuster – im Regelfall KWICs, aber unter Umständen auch größerer Volltextstellen wie in (14) – kann bereits eine erkenntnisleitende Funktion erfüllen und demzufolge einen eigenen Nutzwert haben. Erkenntnisleitende Ord-

nung bedeutet, dass sich die Strukturen von Wortverbindungsmustern, die distinktiven Merkmale ihres Gebrauchs und ihre wechselseitigen Vernetzungen über die linguistische Systematisierung der Kotextmuster in Feldern, abgestuften Hierarchien mit den entsprechenden Verlinkungen und Verweisen, quasi von selbst erschließen. Man kann so bereits auf der rein objektsprachlichen Ebene, also auf der Basis vieler gebündelter realer – nicht idealisierter – sprachlicher Vorkommen mit verwandten Merkmalen, etwas über den sprachlichen Usus erfahren.

Allerdings erfordert die Rezeption derart visualisierter Ergebnisse schon ein gewisses Training im »Lesen von KWICs«. Daher wird die Beschränkung auf diese pure systematisierte Darstellung der Daten selbst wohl nur den wirklichen Spezialisten vorbehalten bleiben. Anzustreben wäre trotzdem, diese Darstellungsebene auch zugänglich zu machen, was nicht zuletzt eine neue Qualität in der Nachvollziehbarkeit empirischer Schlüsse bedeuten würde. Sollen die korpusgesteuert erarbeiteten Ergebnisse darüber hinaus auch als Anschlussstelle für Weiterverarbeitungen sowohl in theoretischer Hinsicht als auch für praxisnahe Anwendungen dienen, kann auch eine so datennahe linguistische Interpretation nicht auf metasprachliche Kommentierungen verzichten. In diesem Sinne sind auch die Kategorien der Abbildung 2 als metasprachliche Kommentare zu verstehen, wenn auch stark komprimiert als kategoriale Annotation. Es handelt sich lediglich um ein graduelles Phänomen der Explizitheit.

Unabhängig von der Form der Beschreibung, ihrer Explizitheit und dem Grad des vorausgesetzten Expertenwissens reflektieren auf korpusgesteuerten Analysen basierende Kommentare aber immer den typischen Gebrauch der Wortverbindungen und ihre wechselseitigen Vernetzungen und Hierarchien im Korpus. Dies betrifft alle Ebenen, nicht nur Bedeutung und Stilistik im engen Sinne, sondern auch die Typik des grammatischen Gebrauchs, der Variation usw.

Aus induktiv-datengeleiteter Sicht gibt es keinen Unterschied zwischen formal-strukturellen und grammatischen Angaben im Sinne von Phrasenmuster- oder Restriktionsangaben einerseits und semantischen, stilistischen und kommunikativ-pragmatischen Informationen andererseits (vgl. Jesenšek 2006). So stellen die Kommentare zu Wortverbindungsmustern auf verschiedenen Abstraktionsebenen in (4a) – (6a) und die Funktionsbeschreibungen in den Beispielen (7) – (10) ebenso Gebrauchskommentare dar wie klassische Bedeutungserläuterungen in Wörterbuchartikeln. Die zentralen Aspekte der Gebrauchstypik, die im Mittelpunkt korpusgesteuerter Beschreibungen usueller Wortverbindungen stehen, lassen sich noch einmal wie folgt zusammenfassen:

- Lexikalisierungsgrad/grammatische Fixiertheit (Invarianz/Varianz)
- Kommunikative Einheit (Bedeutung&Funktion)
- Externe Kotextmuster (Kontextmarker)
- Beziehungen zu anderen Wortverbindungen (WV-Netze, WV-Gruppen)
- Hierarchische Bildungsmuster (Wortverbindungsmuster, Lücken und Platzhalter, Funktion des Bildungsmusters)

Für linguistische Experten und Multiplikatoren des Deutschen kann – wie gesagt – eine Darstellung, die vor allem die Daten und ihre Originaltextausschnitte aus den Korpora selbst sprechen lässt, sehr hilfreich sein. Diese Beschreibungsform impliziert ein hohes Maß an abstrahierender linguistischer Reflexion, die aber nicht in jedem Falle sichtbar sein muss. Es geht um den Zugang zu möglichst umfangreichem, am aktuellen sprachlichen Usus orientiertem Datenmaterial, was Nutzer auf diesem gehobenen Level der Sprachbe-

herrschaft immer wieder fordern. Gleichzeitig kommen hier die Stärken solcher Clusterverfahren wie die Kookkurrenzanalyse vor allem in Bezug auf eine feine Ausdifferenzierung von Verwendungskontexten und Gebrauchsbedingungen besonders gut zum Tragen. Für Deutschlerner erweisen sich zu viele »ungeschminkte« Korpusinformationen dagegen eher als kontraproduktiv. Und es ist daher durchaus sinnvoll, auch Korpusdaten in ein – wenn auch modernes – didaktisches Gewand zu hüllen und entsprechend zu konturieren.

5. Anwendungsszenarien

Eine korpusgesteuerte, datennahe Basisempirie, wie sie eben skizziert wurde, ist also das Eine. Die Erfordernisse praxisnaher Anwendungen sind das Andere. Lexikografen und Fremdsprachendidaktiker z.B. müssen eindeutige und oft genug auch ganz pragmatische Entscheidungen treffen, Auswahlkriterien im Vorhinein bestimmen, Informationen bündeln, für verschiedene Nutzergruppen mit unterschiedlichen Kompetenzgraden in unterschiedlichen Produktions- und Rezeptionssituationen lehr- und lernbar machen. Dabei sollen ihre Ergebnisse gleichzeitig der Dynamik und Flexibilität des aktuellen Sprachgebrauchs gerecht werden. Deshalb sind immer wieder Vorschläge unterbreitet worden, die die jeweilige Nutzungssituation ins Zentrum für lexikografische Konzeptionen rücken (z.B. Engelberg/Lemnitzer 2001: 67ff.).¹⁷

Vor diesem Hintergrund ist zu fragen, wie das in diesem Beitrag vorgeschlagene Beschreibungsmodell usueller Wortverbindungen für praxisnahe Bereiche und für den lexikografischen Alltag anwendbar sein könnte. Zunächst stellt es ein wichtiges empirisches Fundament dar, das sowohl in seiner Ganzheit und Offenheit als auch in seiner feinen Ausdifferenzierung der Kontexte und Restriktionen für die linguistische Forschung und Theorienbildung sowie für anwendungsbezogene Weiterverarbeitungen – wenn auch über mehrere Transferstufen hinweg – von Nutzen sein kann. Ein besonderer Erkenntniswert liegt darin, dass mit Hilfe automatischer Clusterverfahren hochgradig differenzierte kontextsensitive Gebrauchsnuancen in neuer Dimension und Tiefe erkannt werden können, deren Kenntnis und Beherrschung – wie bereits betont – gerade für das mittlere bis gehobene Sprachniveau unverzichtbar sind. Wenn man diese Analysen jedoch entsprechend tiefgründig und detailliert vornimmt, kommt man über die Beschreibung kleiner Sprachausschnitte kaum hinaus. »Nützlich« im Sinne der Praxis sind aber eigentlich erst aufbereitete Sprachdaten in einer gewissen Menge. Es handelt sich fast um eine Quadratur des Kreises: möglichst viele Daten in möglichst großer Komplexität, aber bei gleichzeitiger Rekonstruktion der Feinheiten des distinktiven Gebrauchs sprachlicher Einheiten. Dies ist ein Konzept für eine Langzeitforschung. Es muss jedoch möglich sein, in realistischen Abständen eine eng am sprachlichen Usus orientierte Beschreibung usueller Wortverbindungen und eine entsprechende Verbesserung praxisnaher Produkte zu erreichen. Aus diesem Grund plädiere ich dafür, verschiedene korpusmethodische Wege parallel und miteinander verzahnt zu verfolgen.

¹⁷ Zur Reflexion und Systematisierung verschiedener Ebenen der Erarbeitung automatischer und/oder lexikografischer Sprachdaten vgl. ausführlich Müller-Spitzer (2003).

Auf dem hier skizzierten Weg, der korpusgesteuerten Suche nach rekurrenten Bildungsmustern und Strukturen von Wortverbindungen und ihrer systematischen Erklärungen, stehen wir noch am Anfang. Deshalb ist die gleichzeitige Verifizierung bereits erarbeiteten linguistischen Wissens und der entsprechenden Ressourcen (z.B. bereits kodifizierter Wortverbindungsinventaren in Wörterbüchern und phraseologischen Sammlungen) mit Hilfe der neuesten Korpusanalysemethoden und anhand der größten Sprachausschnitte, die uns zur Verfügung stehen, ebenso nützlich und Erfolg versprechend.¹⁸ Wenn sich beide methodische Herangehensweisen wechselseitig befruchten, können auch in Hinblick auf eine moderne Mehrwortlexikografie beträchtliche Fortschritte erreicht werden. Dies betrifft z.B.:

- Die Erarbeitung von Inventaren lexikalisierte useller Wortverbindungen des Deutschen.¹⁹ Solche dokumentarischen Produkte sind vor allem auf die Dokumentation von mehr oder weniger fest lexikalisierten Wortverbindungen auf der Basis von Korpora ausgerichtet. Dabei werden bereits kodifizierte Sammlungen/Listen von Wortverbindungen mit Hilfe korpusanalytischer Verfahren validiert und neue Kandidaten auf diesem Weg ermittelt. Solche Inventare können zwar nicht alle vorstellbaren festen Wortverbindungen, die potenziell zum Bestand einer Sprachgemeinschaft gehören, enthalten. Aber die, die korpusanalytisch ermittelt und validiert wurden, weisen eine gesicherte Usualität auf und können als Referenzinventare für Wörterbücher und Lehrwerke dienen. In diesem Sinne bereits entwickelte Vorstellungen gibt es vor allem auf dem Gebiet der Sprichwortforschung mit der Forderung nach einem »Sprichwörter-Optimum« einer Sprache (Đurčo 2005: 90);
- Die Erarbeitung spezieller lexikografischer und didaktischer Produkte;
- Lehrmaterialien orientiert an Fremdsprachenkompetenzstufen;
- Beschreibungen von Wortschatzausschnitten/Feldern; z.B. zu kulturellen Symbolen wie Körperteilbezeichnungen, Zahlen, Tiernamen usw., zu einzelnen Diskursausschnitten und zu diachronen Schnitten (»neue«/»alte« Wortverbindungen);
- Beschreibungen von Mehrwortklassen (z.B. Sprichwörter; kommunikative Formeln, Wortpaare, Zurückweisungshandlungen);
- Übersetzungshilfen, für die besonders die feine Kontextsensitivität solcher automatischer Methoden wie die Kookkurrenzanalyse von Nutzen sein können.

Zwischen allen Anwendungsbereichen bestehen natürlich Schnittstellen und fließende Übergänge. So können konstruktionsorientierte Forschungsergebnisse wie die im 3. Punkt gezeigten Beispiele für Wortbindungsmuster und ihre Schemafunktionen auch in konkrete Anwendungsprodukte einfließen (z.B. bei der Lemmaentscheidung, der Angabe von obligatorischen und fakultativen Elementen, funktionalen Restriktionen) usw. Sie können aber auch in Konstruktions- und Kollokationswörterbüchern verarbeitet werden (Engelberg/Lemnitzer 2001: 29ff., Steyer 2008). Ein Konstruktionswörterbuch, das die korpusanalytisch ermittelten und validierten rekurrenten Bildungsmuster von Wortverbindungen und

¹⁸ Die aktuellsten Sammelbände auf dem Gebiet der Phraseologie bieten einen guten Überblick über die wachsende Vielfalt korpusbasierter Forschungen und Projekte bzw. den massiv betriebenen Entwicklungen im sprachtechnologischen Bereich, z.B. bei der Erstellung lexikalischer Ressourcen wie Kollokations- und Idiomdatenbanken und der Entwicklung der entsprechenden Tools (Häcki Buhofer/Burger 2006, Hallsteinsdóttir/Farø 2006).

¹⁹ Einen solchen Weg beschreitet z.B. das Projekt »Kontrastive Kollokationsforschung Deutsch-Slowakisch« (vgl. Đurčo 2008).

ihre typischen Realisierungen in systematisierter Form verarbeitet, wäre sicherlich ein großes Ziel für die Zukunft der Mehrwortlexikografie.

Das in diesem Beitrag skizzierte Beschreibungsmodell ist aus unseren Forschungsarbeiten der letzten Jahre erwachsen. Es basiert auf systematischen empirischen Analysen und umfangreichen Tests, auch in Hinblick auf eine praxisnahe Umsetzung.²⁰ Sehr deutlich ist dabei geworden, dass noch viel empirische Grundlagenforschung im Bereich der korpusgesteuerten Beschreibung usueller Wortverbindungen zu leisten ist und dass es gilt, anwendungsfähige, prototypische Modelle zu entwickeln, die dann wieder Rückwirkungen auf praxisnahe Anwendungsgebiete haben können.

6. Literatur

(a) Wörterbücher

GWDS (2000) = *Duden – Das große Wörterbuch der deutschen Sprache* (PC-Bibliothek Version 2.01).

(b) Sonstige Literatur

Belica, Cyril (1995): *Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analyse-methode*. – Mannheim: Institut für Deutsche Sprache [COSMAS II: <https://cosmas2.ids-mannheim.de/cosmas2-web/>].

Belica, Cyril/Steyer, Kathrin (2008): „Korpusanalytische Zugänge zu sprachlichem Usus.“ – In: M. Vachková (Hg.): *Beiträge zur bilingualen Lexikographie*, 7-24. Tiskárna: Reprošředisko MFF UK, Praha.

Blumenthal, Peter (2006): *Wortprofil im Französischen*. – Tübingen: Niemeyer (= *Beihefte zur Zeitschrift für Romanische Philologie* 332).

Brunner, Annelen/Steyer, Kathrin (2007a): „Corpus-driven study of multi-word expressions based on collocations from a very large corpus.“ – In: *Proceedings of the 4th Corpus Linguistics conference, University of Birmingham, 27–30 July* [http://corpus.bham.ac.uk/corplingproceedings07/paper/182_Paper.pdf].

– (2007b): „Phraseologische und phraseographische Aspekte korpusgesteuerter Empirie.“ – In: V. Jesenšek, M. Fabčić (Hgg.): *Phraseologie kontrastiv und didaktisch. Neue Ansätze in der Fremdsprachenvermittlung*, 181–194. Maribor: Slavistično društvo: Filozofska fakulteta (= *Zora* 47).

Burger, Harald (2007): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 3., neu bearbeitete Auflage. – Berlin: Erich Schmidt.

²⁰ Ergebnisse dieser Pilotarbeiten veröffentlichen wir laufend im Internetauftritt meines Projektes, der Webseite »Wortverbindungen online« [<http://www.ids-mannheim.de/ll/uvwv/wv-online/wv-online.html>] (zuletzt besucht am 29.4.2008). Neben korpusbasierten Kurzstudien findet man dort auch Links zu lexikografischen Artikeln, die von meinem Projekt für das OWID-Portal, das *Online-Wortschatz-Informationssystem Deutsch* am IDS, erarbeitet werden [www.owid.de] (zuletzt besucht am 29.4.2008). Zur lexikografischen Konzeption vgl. Steyer/Lauer/Brunner (2008).

- Burger, Harald/Dobrovol'skij, Dmitrij O./Kühn, Peter/Norrick, Neal R. (Hgg.) (2007): *Phraseologie/Phraseology. Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*. 2 Bde. – Berlin, New York: de Gruyter.
- Burger, Harald/Häcki Buhofer, Annelies/Gréciano, Gertrud (Hgg.) (2003): *Flut von Texten – Vielfalt der Kulturen*. – Baltmannsweiler: Schneider-Verlag Hohengehren.
- Deppermann, Arnulf (2006): „Construction Grammar – Eine Grammatik für die Interaktion?.“ – In: A. Deppermann, R. Fiehler, T. Spranz-Fogasy (Hgg.), 43–65.
- Deppermann, Arnulf/Fiehler, Reinhard/Spranz-Fogasy, Thomas (Hgg.) (2006): *Grammatik und Interaktion. Untersuchungen zum Zusammenhang zwischen grammatischen Strukturen und Gesprächsprozessen*. – Radolfzell: Verlag für Gesprächsforschung.
- Dobrovol'skij, Dmitrij O. (1995): *Kognitive Aspekte der Idiom-Semantik. Studien zum Thesaurus deutscher Idiome*. – Tübingen: Gunter Narr.
- Donalies, Elke (2007): „Marginale Morphologie. Das Problem mit den Rändern – Ein Randproblem?“ – *Sprachreport* 1/2007, 6–12.
- Duffner, Rolf (2006): „Satzadverbien – korpusbasiert.“ – In: A. Näf, R. Duffner (Hgg.), 133–155.
- Durčo, Peter (2005): *Sprichwörter in der Gegenwartssprache*. – Trnava: Univerzita sv. Cyrila a Metoda.
- (2008): „Zum Konzept eines zweisprachigen Kollokationswörterbuchs. Prinzipien der Erstellung. Am Beispiel Deutsch . Slowakisch.“ – *Lexicographica* 24, 69–89 (Thementeil: *Kollokationen in der europäischen Lexikographie und Wörterbuchforschung*).
- Engelberg, Stefan/Lemnitzer, Lothar (2001): *Lexikographie und Wörterbuchbenutzung*. – Tübingen: Stauffenburg.
- Feilke, Helmuth (1996): *Sprache als soziale Gestalt. Ausdruck, Prägung und die Ordnung der sprachlichen Typik*. – Frankfurt a. M: Suhrkamp.
- (2004): „Kontext – Zeichen – Kompetenz. Wortverbindungen unter sprachtheoretischem Aspekt.“ – In: K. Steyer (Hg.), 41–64.
- (2007): „Syntaktische Aspekte der Phraseologie III: *Construction Grammar* und verwandte Ansätze.“ – In: H. Burger, D. Dobrovol'skij, P. Kühn, N. R. Norrick (Hgg.), 63–76. Berlin, New York: de Gruyter.
- Firth, John Rupert (1957): „A Synopsis of Linguistic Theory 1930–1955.“ – In: *Studies in Linguistic Analysis*, 1–32. Oxford: Oxford University Press (reprinted in: Palmer, Frank (Hg.) (1968): *Selected Papers of J. R. Firth 1952–1959*. London, Harlow: Longman).
- Fischer, Kerstin/Stefanowitsch, Anatol (Hgg.) (2006a): *Konstruktionsgrammatik. Von der Anwendung zur Theorie*. – Tübingen: Stauffenburg.
- (2006b): „Konstruktionsgrammatik. Ein Überblick.“ – In: K. Fischer, A. Stefanowitsch (Hgg.): *Konstruktionsgrammatik. Von der Anwendung zur Theorie*, 3–17. – Tübingen: Stauffenburg.
- Fleischer, Wolfgang (²1997): *Phraseologie der deutschen Gegenwartssprache*. – Tübingen: Niemeyer.
- Häcki Buhofer, Annelies/Burger, Harald (Hgg.) (2006): *Phraseology in Motion I. Methoden und Kritik. Akten zur Internationalen Tagung zur Phraseologie, Basel 2004*. – Baltmannsweiler: Schneider-Verlag Hohengehren.
- Hallsteinsdóttir, Erla/Farø, Ken (Hgg.) (2006): *Neue theoretische und methodische Ansätze in der Phraseologieforschung. New Theoretical and Methodological Approaches to Phraseology* (= *Linguistik online* 27, 2/06) [http://www.linguistik-online.de/27_06/].
- Hanks, Patrick (2004): „The Syntagmatics of Metaphor and Idiom.“ – *International Journal of Lexicography* 17/3, 245–274.
- Hausmann, Franz Josef (2004): „Was sind eigentlich Kollokationen?“ – In: K. Steyer (Hg.), 309–334.
- Jesenšek, Vida (2006): „Phraseologie und Fremdsprachenlernen. Zur Problematik einer angemessenen phraseodidaktischen Umsetzung.“ – In: E. Hallsteinsdóttir, K. Farø (Hgg.): *Neue theoretische*

- und methodische Ansätze in der Phraseologieforschung. *New Theoretical and Methodological Approaches to Phraseology*, 137–147.
- Kämper, Heidrun/Eichinger, Ludwig M. (Hgg.) (2007): *Sprach-Perspektiven. Germanistische Linguistik und das Institut für Deutsche Sprache*. – Tübingen: Gunter Narr.
- Keibel, Holger/Belica, Cyril (2007): „CCDB: A Corpus-Linguistic Research and Development Workbench.“ – In: *Proceedings of the 4th Corpus Linguistics conference, Birmingham, 27–30 July* [http://corpus.bham.ac.uk/corplingproceedings07/paper/134_Paper.pdf].
- Klosa, Annette (Hg.) (2008): *Lexikografische Portale im Internet* (= *OPAL* 2008/1, Sonderband) [www.ids-mannheim.de/opal].
- Lemnitzer, Lothar/Zinsmeister, Heike (2006): *Korpuslinguistik : Eine Einführung*. – Tübingen: Gunter Narr.
- Mattheier, Klaus J. (Hg.) (1997): *Norm und Variation*. – Frankfurt a. M. et al.: Peter Lang (= *Forum Angewandte Linguistik* 32).
- Müller-Spitzer, Carolin (2003): „Ord nende Betrachtungen zu elektronischen Wörterbüchern und lexicographischen Prozessen (Classifying considerations related to electronic dictionaries and lexicographic processes; Des considérations classificantes en ce qui concerne des dictionnaires électroniques et des procédés lexicographiques).“ – *Lexicographica* 19, 140–168.
- Näf, Anton/Duffner, Rolf (Hgg.) (2006): „Korpuslinguistik im Zeitalter der Textdatenbanken. Corpus linguistics in the era of text data banks“ (= *Linguistik online* 28, 3/06) [http://www.linguistik-online.de/28_06/].
- Perkuhn, Rainer (2007): „‘Corpus-driven’: Systematische Auswertung automatisch ermittelter sprachlicher Muster.“ – In: H. Kämper, L. M. Eichinger (Hgg.), 465–491.
- Siepmann, Dirk (2002): „Eigenschaften und Formen lexikalischer Kollokationen: Wider ein zu enges Verständnis.“ – *Zeitschrift für französische Sprache und Literatur* 112/1, 240–263.
- Sinclair, John (1991): *Corpus, Concordance, Collocation*. – Oxford: Oxford University Press.
- Steyer, Kathrin (2000): „Usuelle Wortverbindungen des Deutschen. Linguistisches Konzept und lexikografische Möglichkeiten.“ – *Deutsche Sprache* 28/2, 101–125.
- (2003): „Korpus, Statistik, Kookkurrenz. Lässt sich Idiomatisches ‘berechnen’?“ – In: H. Burger, A. Häcki Buhofer, G. Greciano (Hgg.), 33–46.
- (Hg.) (2004a): *Wortverbindungen – mehr oder weniger fest*. – Berlin, New York: de Gruyter (= *Jahrbuch des Instituts für Deutsche Sprache* 2003).
- (2004b): „Kookkurrenz. Korpusmethodik, linguistisches Modell, lexikografische Perspektiven.“ – In: K. Steyer (Hg.), 87–116.
- (2008): „Kollokationen in deutschen Wörterbüchern und in der deutschen Wörterbuchforschung.“ – *Lexicographica* 24 (Thementeil: *Kollokationen in der europäischen Lexikographie und Wörterbuchforschung*), 185–207.
- Steyer, Kathrin/Lauer, Meike (2007): „‘Corpus-Driven’: Linguistische Interpretation von Kookkurrenzbeziehungen.“ – In: H. Kämper, L. M. Eichinger (Hgg.), 493–509.
- Steyer, Kathrin/Lauer, Meike/Brunner, Annalen (2008): „Usuelle Wortverbindungen in OWID. Lexikografische Aspekte unter besonderer Berücksichtigung der Pragmatik.“ – In: A. Klosa (Hg.), 107–121.
- Stubbs, Michael (1997): „Eine Sprache idiomatisch sprechen: Computer, Korpora, Kommunikative Kompetenz und Kultur.“ – In: K. J. Mattheier (Hg.), 151–167.
- Tognini-Bonelli, Elena (2001): *Corpus Linguistics at Work*. – Amsterdam, Philadelphia: Benjamins (= *Studies in Corpus Linguistics* 6).
- Tomasello, Michael (2006): „Konstruktionsgrammatik und früher Erstspracherwerb“ (übersetzt von Stefanie Wulff und Arne Zeschel). – In: K. Fischer, A. Stefanowitsch (Hgg.), 19–37.